

STORAGE CONTROLLER

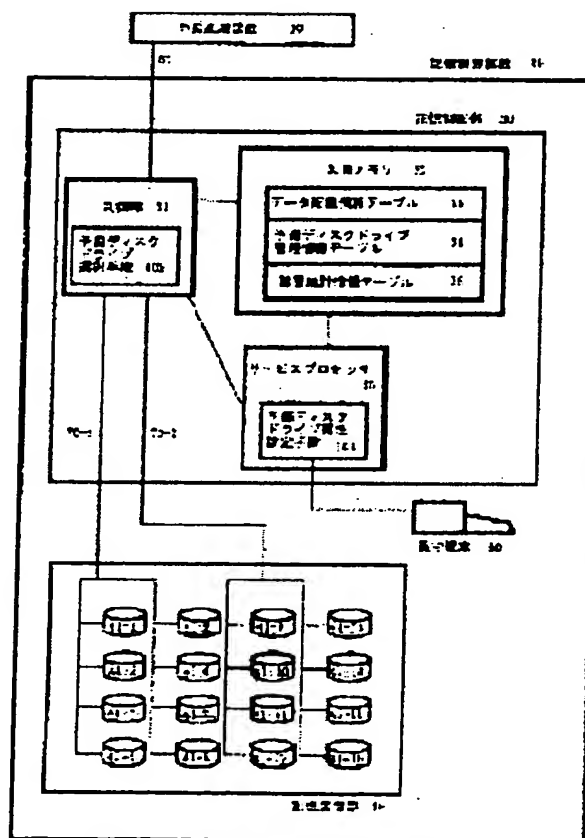
Patent number: JP9330182
 Publication date: 1997-12-22
 Inventor: MATOBA HIROZUMI; SATO TAKAO; FUKUOKA MIKIO;
 OGATA MIKITO
 Applicant: HITACHI LTD
 Classification:
 - international: G06F3/06; G06F3/06; G06F3/06; G06F12/16
 - european:
 Application number: JP19960172986 19960612
 Priority number(s):

Report a data error here

Abstract of JP9330182

PROBLEM TO BE SOLVED: To minimize the cost needed to secure necessary reliability by preferentially using a stand-by disk drive for a disk drive wherein data of relatively high importance are stored and efficiently using the stand-by disk drive when the data are backed up against fault occurrence.

SOLUTION: A table 35 in a common memory 32 contains fault statistic information on disk drives 41-1-41-16. Further, a service processor 36 is controlled by an operator through a maintenance terminal 50 to set necessary information in a data arrangement information table 33 and a stand-by disk drive management table 34 in the common memory 32. Then a stand-by disk drive selecting means 102 of a control part 31 preferentially uses the stand-by disk drive for the disk drive where data of relatively high importance are stored to efficiently use the stand-by disk drive when backing up the data while carrying on data transmission to and reception from a central processor 10 during fault occurrence in a disk drive.



Data supplied from the esp@cenet database - Patent Abstracts of Japan

D4

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Control two or more independent disk drives, control transfer of the data between a central processing unit and said two or more disk drives, and at least one of said two or more disk drives is received. Make said data arrange and memorize and the control logic which processes the access request to said two or more disk drives from said central processing unit to compensate for said data arrangement is included. Recognize the failure generated in either of said two or more disk drives, and it recognizes whether it is the failure generated in which disk drive. It is the memory control unit which backs up the data memorized on the disk drive concerned, without interrupting transfer of said data to the reserve disk drive of said two or more disk drives. At least one of said two or more disk drives is held as a reserve disk drive.

[whether it considers as the common reserve disk drive used to every disk drive in the case of a backup process, and] Or a means to set up beforehand the attribute whether to consider as the reserve disk drive of the dedication used only to one or more specific disk drives in the case of a backup process, The memory control unit characterized by having a selection means to choose one of said reserve disk drives as backup processes with reference to the contents of a setting of the attribute of each reserve disk drive of said in case a backup process is started.

[Claim 2] In a memory control unit according to claim 1 said selection means In choosing one of said reserve disk drives to the disk drive which the failure generated What [one or more] is [a thing intact among said common reserve disk drives] intact among the reserve disk drives of dedication for the disk drive which those or more with one and said failure generated At a certain time The memory control unit characterized by choosing one of intact things among said common reserve disk drives.

[Claim 3] The memory control unit characterized by having further a means to divide said data into two or more segments, to store each segment in a different disk drive among said two or more disk drives in a memory control unit according to claim 1 or 2, to arrange data, and to process the access request to said store from said central processing unit to compensate for said data arrangement.

[Claim 4] The memory control unit characterized by having further a means to divide said data into two or more segments, to create redundancy data further in a memory control unit according to claim 1 or 2 based on them, to store said segment and redundancy data in a different disk drive among said two or more disk drives, to arrange data, and to process the access request to said store from said central processing unit to compensate for said data arrangement.

[Claim 5] In a memory control unit according to claim 4, a failure occurs in one set of the disk drive of two or more disk drives which stored said each segment and redundancy data. When carrying out the backup process to the disk drive which this failure generated, The memory control unit characterized by having further a means to perform a backup process by reconfiguring the data stored on the disk drive which the failure generated using the data stored in disk drives other than the disk drive which the failure generated.

[Claim 6] The memory control unit characterized by having further the means which changes whether the backup process by whether extent of the failure of the disk drive which the failure

generated performs the backup process which depends on reconstruction of the aforementioned data in said backup process in a memory control unit according to claim 5, and copying the data stored on the disk drive which said failure generated simply is performed.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] In the memory control unit with which this invention processes the disk drive group of a redundant configuration especially about a storage control technique Continuing transfer of data with a central processing unit, when a defect or a temporary failure occurs in a disk drive The equipment which maintains high dependability by starting automatically the backup process which used the spare disk drive about the data and redundancy data on the disk drive which the failure generated, Or it sets to the memory control unit which processes an un-redundancy disk drive group. Continuing transfer of data with a central processing unit, when the accumulating totals of a failure temporary about each disk drive exceed a threshold By starting automatically the backup process which used the spare disk drive about the data on the disk drive with which the accumulating totals of the failure exceeded the threshold, it applies to the equipment which maintains high dependability, and is related with an effective technique.

[0002]

[Description of the Prior Art] For the purpose of securing the data integrity accumulated in the field of information processing The dependability of stored data is raised by storing, where the redundancy of data is raised to an independent disk drive physically [plurality]. The disk drive group which detects it when a defect occurs in a disk drive, and is mutually associated logically about generation and storing of redundancy data, Namely, reconfigure the data on the disk drive concerned using the data and redundancy data which are on the disk drive of others other than the disk drive concerned among redundancy groups, and a backup process is automatically started to a spare disk drive. On the other hand, when the data with which the access request from a central processing unit is arranged on the failure disk drive are received, the approach of reconfiguring the data concerned using redundancy data and answering a central processing unit is indicated by Patent Publication Heisei No. 504431 [05 to]. Moreover, with an identical text document, about a reserve disk drive, fix beforehand as an object for backup of a specific redundancy group, and each reserve disk drive is not set up. By considering as the common reserve disk pool used as every object for redundancy groups It supposes that there is no futility to the last one of a reserve disk drive at the time of the need, and it is usable, and the approach of pressing down low is indicated [cost / presser-foot] few in the number of reserve disk drives taken to secure required dependability.

[0003]

[Problem(s) to be Solved by the Invention] In the above-mentioned conventional technique, when two or more groups actuation of the redundancy group which consists of two or more disk drives is carried out, continuing transfer of data with a central processing unit, a failure occurs in the midst to which the backup process is automatically performed about a certain redundancy group, or one of the disk drives in redundancy group another after the processing is completed, and there may be the situation where a backup process is needed. When the approach of a reserve disk pool is used at this time, one intact thing is automatically selected in all reserve disk drives, and a backup process is started. That is, the reserve disk drive will be consumed by the backup process started automatically sequentially from the redundancy group which the

failure generated. Therefore, there may be the situation where the reserve disk drive equipped when the approach of a reserve disk pool is used will be consumed altogether in the inside of a short time.

[0004] Differentiation which depends on the redundancy group which becomes each backup origin in the case of this backup process is not performed. That is, differentiation according to whether data with a low significance are stored relatively or important data are stored relatively, such as data temporary in the redundancy group of a backing up agency, is not performed about use of a reserve disk drive. For this reason, for the backup process of the disk drive with which data with a low significance were stored relatively After using the reserve disk drive of the last which was intact and remained all over the reserve disk pool, When a failure occurs by the redundancy group in which data with a high significance were stored relatively and a backup process is needed A backup process cannot be automatically started about the redundancy group. It is extended until it succeeds in exchange of the failure disk drive by the customer engineer, or extension of a reserve disk drive. Moreover, although there is a problem that must continue actuation with an un-redundancy configuration and data integrity with a high significance falls relatively as a result in order to continue transfer of data with a central processing unit also while the backup process will be carried out to extension Consideration about this point is not carried out.

[0005] [when it is possible for a user or a customer engineer to decide beforehand in which redundancy group data with a high significance are generally relatively stored among the data stored, and to employ it] Those who used the reserve disk drive for backup preferentially to the redundancy group in which important data are stored relatively can make data integrity higher efficiently to a fixed number of numbers of reserve disk drives. However, when a reserve disk drive is beforehand assigned fixed altogether to each redundancy group, efficient use of a reserve disk drive becomes impossible.

[0006] Therefore, it becomes the technical problem of this invention to realize coexistence of pressing down low the cost required in order to secure required dependability to assign a reserve disk drive preferentially to the redundancy group in which the data with which a user judges a backup process that significance is high relatively in the equipment started automatically were stored, and by using the given reserve disk drive efficiently.

[0007]

[Means for Solving the Problem] In order to solve the above-mentioned technical problem, this invention controls two or more independent disk drives. A central processing unit, Control transfer of the data between said two or more disk drives, and at least one of said two or more disk drives is received. Make said data arrange and memorize and the control logic which processes the access request to said two or more disk drives from said central processing unit to compensate for said data arrangement is included. Recognize the failure generated in either of said two or more disk drives, and it recognizes whether it is the failure generated in which disk drive. It is the memory control unit which backs up the data memorized on the disk drive concerned, without interrupting transfer of said data to the reserve disk drive of said two or more disk drives. At least one of said two or more disk drives is held as a reserve disk drive.

[whether it considers as the common reserve disk drive used to every disk drive in the case of a backup process, and] Or a means to set up beforehand the attribute whether to consider as the reserve disk drive of the dedication used only to one or more specific disk drives in the case of a backup process, In case a backup process is started, he is trying to have a selection means to choose one of said reserve disk drives as backup processes with reference to the contents of a setting of the attribute of each reserve disk drive of said.

[0008] Moreover, are in charge of said selection means choosing one of said reserve disk drives to the disk drive which the failure generated. What [one or more] is [a thing intact among said common reserve disk drives] intact among the reserve disk drives of dedication for the disk drive which those or more with one and said failure generated At a certain time He is trying to choose one of intact things among said common reserve disk drives.

[0009] Moreover, said data are divided into two or more segments, and each segment is stored in a different disk drive among said two or more disk drives, and he arranges data, and is trying to

have further a means to process the access request to said store from said central processing unit to compensate for said data arrangement.

[0010] Moreover, said data are divided into two or more segments, and redundancy data are further created based on them, and said segment and redundancy data are stored in a different disk drive among said two or more disk drives, and he arranges data, and is trying to have further a means to process the access request to said store from said central processing unit to compensate for said data arrangement.

[0011] Furthermore, when carrying out the backup process to the disk drive which the failure generated in one set of the disk drive of two or more disk drives which stored said each segment and redundancy data, and this failure generated, he is trying to have further a means perform a backup process, by reconfiguring the data stored on the disk drive which the failure generated using the data stored in disk drives other than the disk drive which the failure generated.

[0012] Furthermore, the memory control unit characterized by having further the means which changes whether the backup process by whether extent of the failure of the disk drive which the failure generated performs the backup process which depends on reconstruction of the aforementioned data in said backup process, and copying the data stored on the disk drive which said failure generated simply is performed.

[0013]

[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained to a detail, referring to a drawing. Drawing 1 is the outline block diagram of the configuration of the information processing system containing the disk subsystem which is the gestalt of 1 operation of this invention. It consists of a central processing unit 10 and a memory control unit 20 containing the storage control section 30 and the storage section 40. A central processing unit 10 and the storage control section 30 are combined through the information-transmission line 60, and the storage control section 30 and the storage section 40 are combined by the information-transmission line 70-1 and the information-transmission line 70-2. In the store section 40, a disk drive 41-1 to 41-8 is connected to the control section 31 in the storage control section 30 by the information-transmission line 70-1, and the disk drive 41-9 to 41-16 is connected to the control section 31 in the storage control section 30 by the information-transmission line 70-2. The storage control section 30 consists of shared memories 32 with the control section 31 which controls transfer of the data between a central processing unit 10 and the store section 40, and a service processor 36.

[0014] A shared memory 32 is referred to by a control section 31 and the service processor 36, and contains the reserve disk drive management information table 34, the data arrangement information table 33 for specifying arrangement of the data to the disk drive 41-1 to 41-16 in the store section 40 as a meaning, and the failure statistical information table 35. The statistical information about the failure of each disk drive 41-1 to 41-16 is stored in the failure statistical information table 35. A control section 31 includes the reserve disk drive selection means 102.

[0015] A service processor 36 includes the reserve disk drive attribute setting means 101. A service processor 36 is controlled by the operator, i.e., the user, or customer engineer which the exterior does not illustrate through a maintenance terminal 50, and sets up information required for the data arrangement information reserve disk drive management information table 33 and 34. Namely, the data arrangement information table 33 is received. A setup of the configuration of each redundancy group and a non-redundancy group, i.e., a setup of which to specify as the configuration disk drive of what position of which redundancy group and a non-redundancy group among disk drives 41-1 to 41-16, [whether it is referred to as RAID5 which consists of the arranging method, for example, six drives for data and one disk drive for parity, and] It is referred to as RAID0 used as the disk drive group which set to RAID1 of one drive for data, and one disk drive for parity, or became independent by one set, or A setup about the classification of the method of arranging **, A setup of how many logical devices to assign in what kind of sequence, respectively is performed to each redundancy group and a non-redundancy group. As opposed to the reserve disk drive management information table 34 Similarly the reserve disk drive attribute setting means 101 is used through a maintenance terminal by the operator. [which / of a disk drive 41-1 to 41-16 / is set as a reserve disk drive, and] Moreover,

[whether it considers as the common reserve disk drive which uses each reserve disk drive also to which redundancy group or a non-redundancy group, and] Or it considers as the reserve disk drive of the dedication used only to a specific redundancy group or a specific non-redundancy group, or is set as dedication to which redundancy group or a non-redundancy group, respectively, or ** is set up.

[0016] When there is read-out to a logical device from a central processing unit 10 or a write request, with the data arrangement information table 33, if a control section 31 is required, it will compute correspondence with the device access address, and each disk drive 41-1 to 41-16 in the store section 40 and the storage region within the disk drive using the reserve disk drive management information table 34, and a control section 31 controls each disk drive. The reserve disk drive management information table 34 is referred to because during a backup process and after processing will be able to access original not a disk drive but backup place disk drive.

[0017] An operator becomes possible [assigning a reserve disk drive preferentially to the data concerned at dedication allocation and by arranging important data from a central processing unit relatively to the logical device of this ** further] about a reserve disk drive at the disk drive group containing at least one or more specific logical devices combining above-mentioned allocation of a logical device and reserve disk drive allocation.

[0018] Generating of the error at the time of the control to each disk drive 41-1 to 41-16 and a report of a failure are sorted out and arranged by the control section 31, and are stored and updated by it by the failure statistical information table 35 in a shared memory 32. When a control section 31 judges that a backup process is required from statistical information at the time of updating, the reserve disk drive selection means 102 is used, one of reserve disk drives is chosen automatically, it is made into a backup place disk drive, and the backup process of the data on the disk drive which the failure generated is started automatically.

[0019] As the approach of backup, it is in the approach by simple copy, or the approach of reconfiguring the data on a failure disk drive using redundancy, and writing in a backup place disk drive and *****. In decision of initiation of this backup process, when it is judged that the accumulating totals of the failure of a specific class exceeded the threshold, before resulting in loss of data, backup to a reserve disk drive may be performed by the approach, i.e., the approach by simple copy, of not using redundancy preclusively. This approach is an approach effective in maintaining data integrity also in any in case there is no redundancy, when redundancy is in the disk drive group of this ** at that time. Moreover, in decision of initiation of a backup process, when it is judged that the accumulating totals of the failure of a specific class exceeded another threshold, and redundancy is in the disk drive group of this ** at the time, backup to a reserve disk drive may be performed by reconfiguring the data on the failure generating disk drive concerned using redundancy.

[0020] In selection of a reserve disk drive, for the disk drive group which the failure generated, the reserve disk drive of dedication, When both common reserve disk drives have an intact reserve disk drive Any one may be chosen from from as a backup place disk drive among the reserve disk drives of common use or dedication. Moreover, give priority and one is chosen from from as a backup place disk drive among common reserve disk drives. You may make it choose one of the reserve disk drives of dedication as a backup place disk in the condition that the intact thing also of one was lost with the common reserve disk drive.

[0021] Drawing 2 is the hard block diagram of the information processing system containing the disk subsystem which is the gestalt of 1 operation of this invention. The control section in the principle explanatory view of drawing 1 corresponds to the part containing a host adapter 231-1 to 231-2, a disk adapter 233-1 to 233-4, cache memory 232-1 to 232-2, the data transfer bus 237-1 to 237-2, and the storage control circles communication bus 236 in drawing 2. The disk drive 242-1 to 242-64 of store circles is a SCSI disk drive. Although each disk adapter 233-1 to 233-4 controls only a specific disk drive box among the disk drive boxes 241-1 to 241-2, the information-transmission line 270-1 to 270-16 which connects each disk drive and the disk adapter 233-1 to 233-4 in a control section is raising responsibility to the storage control section 230 while surely raising those with two line, and failure-proof nature. To the write request of the data from a central processing unit 210, it duplex-izes to two cache memory 232-

1 to 232-2, and buffers in it, and from the variable format which is the data format from a central processing unit 210, it changes into the fixed format for SCSI disk drives, and writes in a disk drive. Therefore, since data are surely accessed through a cache, they prevent loss of write request data by (96 hours), and duplex-izing data as mentioned above. [data cache memory] [un-volatilizing] The data transfer between central processing units 210 is raising the response engine performance while two or more information-transmission lines 260-1 to 260-8 are prepared and it raises failure-proof nature. The shared memory 234-1 to 234-2 is duplex-ized, and it prevents that data arrangement information memory etc. is lost. The data transfer bus 237-1 to 237-2 is duplex-ized, and is raising failure-proof nature and the response engine performance. A host adapter 231-1 to 231-2, a disk adapter 233-1 to 233-4, and ***** can operate to juxtaposition, and are raising the response engine performance. Thus, in control of the data between a central processing unit 210 and a disk drive 242-1 to 242-64, all of the information-transmission line 260-1 to 260-8, a host adapter 231-1 to 231-2, a disk adapter 233-1 to 233-4, the data transfer bus 237-1 to 237-2, cache memory 232-1 to 232-2, the information-transmission line 270-1 - 270-16** are redundancy-ized.

[0022] Drawing 4 shows the general-view Fig. of a memory control unit 220. They are **** of 1.3m, the depth of about 0.8m, and magnitude with a height of about 1.8m. There are the two power-source input sections in equipment, and the resistance over the power failure in the outside of equipment is raised. The disk drive current supply unit section 420-1 which is at the bottom of front among the disk drive current supply unit sections 420-1 to 420-8 by the side of the store section 230 is shown in drawing 5 in more detail. This changes input power into a direct current, including a maximum of eight disk drives 242-1 to 242-8, two DC-power-supply feed zones 421-1 to 421-2 which supply a power source to each part are redundant configurations, respectively, and if either operates, a disk drive 242-1 to 242-8 can work. The disk drive box 241-1 of drawing 2 is constituted from the disk drive current supply unit section 420-1 to 420-4, and the disk drive box 241-2 consists of the disk drive current supply unit sections 420-5 to 420-8.

[0023] Signs that it saw from signs that it saw from the front face of the storage control section 220, and a rear face are shown in drawing 6 and drawing 7. In drawing 7, a host adapter 231-1 to 231-2, a disk adapter 233-1 to 233-4, cache memory 232-1 to 232-2, and a shared memory 234-1 to 234-2 are mounted in the Boolean part box. The DC-power-supply intensive box 322 and two or more DC-power-supply feed zones 323-1 to 323-6 are two redundant configurations too, and are equipped with the breaker box 321-1 to 321-2 to each. The above duplex-izing and the inside of the part multiplexed, A shared memory 234-1 to 234-2, cache memory 232-1 to 232-2, a host adapter 231-1 to 231-2, a disk adapter 233-1 to 233-4, a disk drive 242-1 to 242-64, the DC-power-supply feed zone 323-1 to 323-6 And each DC-power-supply feed zone in 421-1 - 421 -16 and the DC-power-supply intensive boxes 322 In addition to the maintenance at the time of failure generating, ***** of scheduled maintenance and a hard configuration is possible, being ***** and exchangeable in hot swapping, and continuing transfer of the data between central processing units.

[0024] Next, some examples are shown and selection of the reserve disk drive at the time of processing and the backup process of the access request from a central processing unit 10 is explained in the principle explanatory view of drawing 1. Drawing 8 shows an example of a logical setup of each disk drive in the store section 40. A disk drive 41-1 to 41-16 by initial setting by the operator Four reserve disk drives (disk drive 41-13 to 41-16) are set up. The remaining disk drives 41-1 to 41-12 It is logically divided into the disk drive group 90-1 to 90-4 which consists of a disk drive 41-1 to 41-3, 41-4 to 41-6, 41-7 to 41-9, and 41-10 to 41-12, respectively. the disk drive which can be mounted in this example -- it all mounted and all they are set up as a disk drive which stores a reserve disk drive or data, and redundancy data. By each disk drive group, data and redundancy data are set as two disk drives and the redundancy disk drive group of RAID5 which consists of three disk drives which occupies one set in capacity, respectively. A disk drive 41-13 is set as 90 to disk drive group 1 dedication among reserve disk drives, and the disk drive 41-14 is set as 90 to disk drive group 2 dedication. Moreover, the disk drive 41-15 and the disk drive 41-16 are set up as a common reserve disk drive.

[0025] In a configuration like drawing 8, the disk drive group 90-1 and 90-2 can perform the backup process to 3 times by max once [at least] in the viewpoint of allocation of a reserve disk drive. That is, if one set of a reserve disk drive is used once, use beyond it cannot be performed (however, failure recovery of the disk drive which the failure generated is performed, and when the data stored in the reserve disk drive are reverse-copied to the disk drive by which failure recovery was carried out, the reserve disk drive can be used again). Therefore, since a common reserve disk drive can be used a maximum of twice and the reserve disk drive of dedication can be used once, when two sets of common reserve disk drives are used by other disk drive groups, it is possible [it is possible to perform the backup process to 3 times at convenience and max, and] to perform at least one backup process. On the other hand, since the reserve disk drive of dedication is not set up, the disk drive group 90-3 and 90-4 cannot be performed once at the maximum, when backup process but another possible [the backup process by the common reserve disk drive to 2 times] has started in advance.

[0026] The logical condition in a certain time in a setup like drawing 8 is shown in drawing 9. Table 1 of drawing 12 shows some contents on the reserve disk drive management information table 34 at this time. Each item in Table 1 is the following semantics. A reserve disk drive location number is a number which shows the arrangement location of each disk drive in the store section 40 of a reserve disk drive. The exclusive attribute flag means having set it as the attribute used for dedication to a specific disk drive group in the case of a backup process, when it is ON. A busy condition expresses [under simple backup, a backup settled, under the backup which used redundancy, and a reverse copy, and] whether it is intact. A thing table is carried out, and when [OFF and whose a busy condition an exclusive attribute flag is except an intact condition] being set up so that it may be used for dedication to the disk drive group to which the disk drive group serial number points when an exclusive attribute flag is ON, a correspondence disk drive group serial number means whether it is using it to the disk drive group to which which disk drive group serial number points. In 01, 02 corresponds to the disk drive group 90-2, and 03 corresponds to the disk drive group 90-1 at the disk drive group 90-3. The relative position of a processing-object disk drive expresses the relative position in the disk drive group of the disk drive for a backup process. For example, 2 of the best column of Table 1 corresponds to the disk drive 41-2 of the disk drive group 90-1.

[0027] Drawing 3 shows the flow chart in case the reserve disk drive selection means 102 chooses a reserve disk drive with reference to the reserve disk drive management information table 34. In drawing 9, it is shown that the failure which needs a backup process for the disk drive to which x mark is attached has occurred, and the arrow head is pointing to the backup place disk drive. Suppose that it is the disk drive with obstacles with which the disk drive 41-12 was most judged to be the backup process need in recently among the disk drives to which x mark is attached. In order that one may not have an intact thing in a common reserve disk drive at this time, the backup process of a disk drive 41-12 is not performed. That is, in the flow chart of drawing 3, it becomes the flow of NO and step 1007 "it is to processing at the time of backup process improper" to NO and step 1004 "the reserve disk drive of dedication is set as the disk drive group concerned" to step 1001 "backup process demand generating" and step 1002 "there is an intact thing among the reserve disk drives which are not dedication." Therefore, access to a disk drive 41-12 is henceforth performed using the redundancy of a disk drive group. Moreover, henceforth, if the disk drive 41-14 which is the last intact reserve disk is not in either of the disk drives about either of the disk drives 41-4 to 41-6 which constitute the disk drive group 90-2 which is set up as an exclusive reserve disk drive when the failure which needs a backup process occurs, a backup process will not be performed. When this backup process operates, it sets to the flow chart of drawing 3. As opposed to step 1001 "backup process demand generating" and step 1002 "there is an intact thing among the reserve disk drives which are not dedication" NO, As opposed to step 1004 "the reserve disk drive of dedication is set as the disk drive group concerned" YES, It becomes the flow of YES, step 1006 "specify one for backup processes from the intact thing which is the reserve disk drive of dedication", and step 1008 "backup process initiation" to step 1005 "there is an intact thing among the reserve disk drives of dedication." However, when extension of a reserve disk drive was performed by the customer

engineer, or when processing which the exchange or repair of a failure disk which finished the backup process is performed, and reverse-copies the contents of the backup place disk drive is performed, the backup process of any disk drive group can become again possible.

[0028] Drawing 10 shows another example of a logical setup of each disk drive in the store section 40. A disk drive 41-1 to 41-16 by initial setting by the operator Four reserve disk drives (disk drive 41-13 to 41-16) are set up. A disk drive 41-1 to 41-10 among the remaining disk drives 41-1 to 41-12 It is logically divided into the disk drive group 90-5 to 90-9 which consists of a disk drive 41-1 to 41-3, 41-4 to 41-6, 41-7 to 41-8, 41-9, and 41-10, respectively. In this example, from the upper limit of the number of disk drives which can be mounted, it was mounted few and two all are set up about a mounted part as a disk drive which stores a reserve disk drive or data, and redundancy data. In the disk drive group 90-5 and 90-6, data and redundancy data are set as two disk drives and the redundancy disk drive group of RAID5 which consists of three disk drives which occupies one set in capacity, respectively. By the disk drive group 90-7, data and redundancy data are set as the redundancy disk drive group of RAID1 which consists of two disk drives which occupies every one disk drive in capacity, respectively. The disk drive group 90-8 and 90-9 are set as the non-redundancy disk drive group of RAID0 which constitutes the disk drive group which one set only of a disk drive became independent of, respectively. A disk drive 41-13 is set as 90 to disk drive group 5 dedication among reserve disk drives, the disk drive 41-14 is set as 90 to disk drive group 6 dedication, and the disk drive 41-15 and the disk drive 41-16 are set up as a common reserve disk drive.

[0029] In a configuration like drawing 10 the disk drive group 90-5 and 90-6 Although it is possible in the viewpoint of allocation of a reserve disk drive to perform the backup process to 3 times by max once [at least] and the backup process to 2 times is possible for the disk drive group 90-7 and 90-8, and 90-9 at the maximum on the other hand When another backup process has started in advance, it cannot carry out once.

[0030] The logical condition in a certain time in a setup like drawing 10 is shown in drawing 11 . Table 2 of drawing 12 shows some contents on the reserve disk drive management information table 34 at this time. In Table 2, in 01 of a correspondence disk drive group serial number, 02 corresponds to 90-6 and 04 corresponds to 90-8 at the disk drive group 90-5.

[0031] In drawing 11 , it is shown that the failure which needs a backup process for the disk drive to which x mark is attached has occurred, and the arrow head is pointing to the backup place disk drive. Suppose that it is the disk drive with obstacles with which the disk drive 41-10 was most judged to be the backup process need in recently among the disk drives to which x mark is attached. In order that one may not have an intact thing in a common reserve disk drive at this time, the backup process of a disk drive 41-10 is not performed. That is, in the flow chart of drawing 3 , it becomes the flow of NO and step 1007 "it is to processing at the time of backup process improper" to NO and step 1004 "the reserve disk drive of dedication is set as the disk drive group concerned" to step 1001 "backup process demand generating" and step 1002 "there is an intact thing among the reserve disk drives which are not dedication."

[0032] Since the disk drive 41-10 belongs to the un-redundancy disk drive group 90-9, if the decision which needs a backup process aims at prophylactic backup, access to a disk drive 41-10 will be henceforth performed, while possibility that some data will lose has been in a high condition. Moreover, henceforth, if the disk drive 41-13 which is the last intact reserve disk is not in either of the disk drives about the disk drive 41-1 to 41-3 which constitutes the disk drive group 90-5 set up as an exclusive reserve disk drive, and either of 41-15 when the failure which needs a backup process occurs, a backup process will not be performed. Namely, if the backup process to 41-3 is completed normally If it becomes before 41-1, 41-2, and 41-15 constitute the disk drive group 90-5 and are completed normally on the other hand, since 41-1, 41-2, and 41-3 constitute the disk drive group 90-5 In the former, the backup process using 41-13 is possible about 41-1, 41-2, and 41-15, and possible about 41-1, 41-2, and 41-3 in the latter. Thus, if the backup process to a disk drive 41-15 is completed about the disk drive 41-3, also when a failure occurs in a disk drive 41-15, the backup process about a disk drive 41-15 is possible.

[0033] When the backup process about the disk drive group 90-5 operates In the flow chart of

drawing 3 Step 1001 "backup process demand generating", As opposed to step 1002 "there is an intact thing among the reserve disk drives which are not dedication" NO, As opposed to step 1004 "the reserve disk drive of dedication is set as the disk drive group concerned" YES, It becomes the flow of YES, step 1006 "specify one for backup processes from the intact thing which is the reserve disk drive of dedication", and step 1008 "backup process initiation" to step 1005 "there is an intact thing among the reserve disk drives of dedication." However, when extension of a reserve disk drive was performed by the customer engineer, or when processing which the exchange or repair of a failure disk which finished the backup process is performed, and reverse-copies the contents of the backup place disk drive is performed, the backup process of any disk drive group can become again possible.

[0034] In the latter of the above two examples of a configuration, although what consists of only disk drives single as an example of an un-redundancy disk drive group was explained, RAID5 without the non-redundancy group which consists of two or more disk drives, for example, parity data, etc. is sufficient as this.

[0035] Thus, a backup process is preferentially performed to the disk drive group with which the reserve disk drive of beforehand dedication was set up when a backup process was equally performed to all the disk drive group while the intact reserve disk drive had allowances in the number as a whole according to the gestalt of this operation, and allowances had been lost in the number as a whole on the other hand. Namely, two or more reserve disk drives given while the intact reserve disk drive had allowances in the number as a whole can be used efficiently. On the other hand, when allowances have been lost in the number as a whole As opposed to the disk drive group which the operator judged beforehand that data with a high significance were stored relatively, and set the reserve disk drive as dedication The data integrity on the disk drive group is preferentially raised to backup disks using a reserve disk drive. The cost required in order to secure the dependability demanded by these is minimized.

[0036] As mentioned above, although this invention was concretely explained based on the gestalt of operation, this invention is not limited to said example and can be variously changed in the range which does not deviate from the summary.

[0037]

[Effect of the Invention] In the equipment which according to the memory control unit of this invention starts a backup process automatically while continuing transfer of data with a central processing unit at the time of failure generating of a disk drive Making it possible to use a reserve disk drive preferentially by setting actuation of a user, a customer engineer, etc. to the disk drive with which data with a high significance were stored relatively The effectiveness of minimizing the cost required in order to secure required dependability by using the given reserve disk drive efficiently is acquired.

[Translation done.]

特開平9-330182

(43) 公開日 平成9年(1997)12月22日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	3 0 6		G 0 6 F 3/06	3 0 6 H
	3 0 4			3 0 4 F
	5 4 0			5 4 0
12/16	3 1 0	7623-5B	12/16	3 1 0 P

審査請求 未請求 請求項の数 6 F D (全 14 頁)

(21) 出願番号 特願平8-172986

(22) 出願日 平成8年(1996)6月12日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 的場 宏純

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 佐藤 孝夫

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 福岡 幹夫

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 笹岡 茂 (外1名)

最終頁に続く

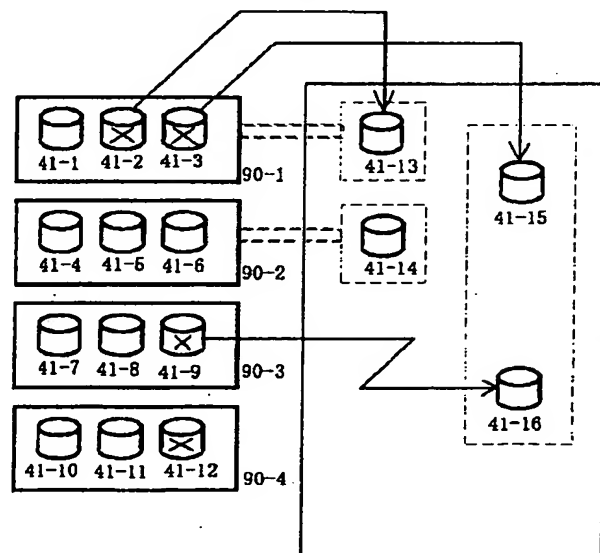
(54) 【発明の名称】 記憶制御装置

(57) 【要約】

【課題】 バックアップ処理において、ユーザが重要度が高いと判断するデータの格納された冗長群に対し予備ディスクドライブを優先的に割当てて。

【解決手段】 図において、90-1～90-4は夫々3台のディスクドライブからなるドライブ群であり、41-13～41-16は予備ドライブである。予備の41-13は群90-1の専用予備であり、予備の41-14は群90-2の専用予備であり、41-15と41-16は共用予備である。まず、41-3が障害となれば、専用予備の41-13より優先して共用予備の41-15が割り当てられ、次に、41-9が障害となれば、共用予備の41-16が残っているので41-16が割り当てられ、次に41-2が障害となれば、共用予備は残っていないので、群90-1用の専用予備の41-13が割り当てられ、次に、41-12が障害となれば、共用予備は残っておらず、41-14は群90-2の専用予備であるため、割当て不可であり、障害ディスク41-12はバックアップされない。

図 9



【特許請求の範囲】

【請求項1】 複数の独立なディスクドライブを制御し、中央処理装置と、前記複数のディスクドライブとの間におけるデータの授受を制御し、前記複数のディスクドライブの内少なくとも1つに対して、前記データを配置し記憶させ、前記中央処理装置からの前記複数のディスクドライブに対するアクセス要求を前記データ配置に合わせて処理する制御論理を含み、前記複数のディスクドライブのいずれかに発生した障害を認識し、それがどのディスクドライブに発生した障害なのかを認識し、前記データの授受を中断すること無しに当該ディスクドライブ上に記憶されているデータを前記複数のディスクドライブの内の予備ディスクドライブへバックアップを行う記憶制御装置であって、

前記複数のディスクドライブの内少なくとも1つを、予備ディスクドライブとして保持し、バックアップ処理の際にどのディスクドライブに対しても使用する共用の予備ディスクドライブとするか、あるいはバックアップ処理の際に特定の1つ以上のディスクドライブに対してのみ使用する専用の予備ディスクドライブとするか、という属性を予め設定する手段と、

バックアップ処理を開始する際に、個々の前記予備ディスクドライブの属性の設定内容を参照して前記予備ディスクドライブの内の1つをバックアップ処理用を選択する選択手段を備えることを特徴とする記憶制御装置。

【請求項2】 請求項1記載の記憶制御装置において、前記選択手段は、障害の発生したディスクドライブに対して前記予備ディスクドライブの内の1つを選択するにあたって、前記共用の予備ディスクドライブの内未使用のものが1つ以上あり、かつ前記障害の発生したディスクドライブのための専用の予備ディスクドライブの内未使用のものが1つ以上あるとき、前記共用の予備ディスクドライブの内未使用のものの内の1つを選択することを特徴とする記憶制御装置。

【請求項3】 請求項1または請求項2記載の記憶制御装置において、

前記データを複数のセグメントに分け、各セグメントを前記複数のディスクドライブの内、異なるディスクドライブに格納してデータを配置し、前記中央処理装置からの前記記憶装置に対するアクセス要求を前記データ配置に合わせて処理する手段を更に備えることを特徴とする記憶制御装置。

【請求項4】 請求項1または請求項2記載の記憶制御装置において、

前記データを複数のセグメントに分け、更にそれらを元に冗長データを作成し、前記セグメントと冗長データを前記複数のディスクドライブの内、異なるディスクドライブに格納してデータを配置し、前記中央処理装置からの前記記憶装置に対するアクセス要求を前記データ配置に合わせて処理する手段を更に備えることを特徴とする

記憶制御装置。

【請求項5】 請求項4記載の記憶制御装置において、前記各セグメントと冗長データを格納した複数のディスクドライブの内の1台のディスクドライブに障害が発生し、該障害が発生したディスクドライブに対するバックアップ処理をするとき、障害の発生したディスクドライブ以外のディスクドライブに格納されたデータを用いて、障害の発生したディスクドライブ上に格納されていたデータの再構成をすることにより、バックアップ処理を行う手段を更に備えることを特徴とする記憶制御装置。

【請求項6】 請求項5記載の記憶制御装置において、前記バックアップ処理において、障害が発生したディスクドライブの障害の程度により、前記のデータの再構成に依るバックアップ処理を行うか、単純に前記障害の発生したディスクドライブ上に格納されているデータを複写することによるバックアップ処理を行うかを切替える手段を更に備えることを特徴とする記憶制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、記憶制御技術に関し、特に、冗長な構成のディスクドライブ群を処理する記憶制御装置において、ディスクドライブに欠陥あるいは一時的な障害の発生した際に、中央処理装置とのデータの授受を続行しながら、その障害の発生したディスクドライブ上のデータ及び冗長データについて予備のディスクドライブを用いたバックアップ処理を自動的に開始することにより高い信頼性を維持する装置、または、非冗長なディスクドライブ群を処理する記憶制御装置において、それぞれのディスクドライブについて一時的な障害の累計が閾値を越えた際に、中央処理装置とのデータの授受を続行しながら、その障害の累計が閾値を越えたディスクドライブ上のデータについて予備のディスクドライブを用いたバックアップ処理を自動的に開始することにより高い信頼性を維持する装置に適用して有効な技術に関する。

【0002】

【従来の技術】情報処理の分野において、蓄積されるデータの保全性を確保する等の目的で、複数の物理的に独立なディスクドライブにデータの冗長度を高めた状態で格納することにより記憶データの信頼性を向上させ、ディスクドライブに欠陥が発生した際にそれを検出し、冗長データの生成及び格納に関し互いに論理的に関連付けられているディスクドライブ群、すなわち冗長群の内、当該ディスクドライブ以外の他のディスクドライブ上にあるデータ及び冗長データを用いて当該ディスクドライブ上のデータを再構成し予備のディスクドライブへバックアップ処理を自動的に開始し、一方、中央処理装置からのアクセス要求がその障害ディスクドライブ上に配置されているデータに対してあった場合は、冗長データを

用いて当該データを再構成し中央処理装置に応答する方法が、特表平〇5-504431号で開示されている。また、同文書では、予備ディスクドライブについて、個々の予備ディスクドライブを特定の冗長群のバックアップ用として予め固定して設定しておくのではなく、どの冗長群用としても使用する共用の予備ディスクプールとすることにより、必要時には予備ディスクドライブの最後の1個まで無駄なく使用可能とし、必要な信頼性を確保するのに要する予備ディスクドライブ数を少なく抑えコストを低く抑える方法が、開示されている。

【0003】

【発明が解決しようとする課題】上記従来技術において、複数のディスクドライブから成る冗長群を複数群動作させていた場合、中央処理装置とのデータの授受を続行しながら、ある冗長群について自動的にバックアップ処理が行われている最中に、あるいは、その処理が完了した後に、別の冗長群の中のディスクドライブの1つに障害が発生し、バックアップ処理が必要になるという事態があり得る。このとき、予備ディスクプールの方法を用いている場合は、自動的に全予備ディスクドライブの中で未使用のものを1つ選び出し、バックアップ処理を開始する。すなわち、自動的に開始されるバックアップ処理により、障害が発生した冗長群から順に予備ディスクドライブを消費していくことになる。よって、予備ディスクプールの方法を用いている場合においても、装備されている予備ディスクドライブが短時間のうちに全て消費されてしまうという事態があり得る。

【0004】このバックアップ処理の際に、それぞれのバックアップ元となる冗長群に依る差別化は行われなない。すなわち、予備ディスクドライブの使用に関して、バックアップ元の冗長群にテンポラリのデータ等の相対的に重要度の低いデータが格納されているか、あるいは相対的に重要なデータが格納されているかによる差別化は行われなない。このため、もし相対的に重要度の低いデータの格納されたディスクドライブのバックアップ処理のために、予備ディスクプール中で未使用で残っていた最後の予備ディスクドライブを使用した後、相対的に重要度の高いデータの格納された冗長群で障害が発生しバックアップ処理が必要となったときには、その冗長群について自動的にバックアップ処理を開始することが出来ず、保守員による障害ディスクドライブの交換、または予備ディスクドライブの増設が為されるまで先延ばしされ、また、バックアップ処理が先延ばしにされている間も、中央処理装置とのデータの授受を続行するために、非冗長な構成のままで動作を続けなければならず、結果として相対的に重要度の高いデータの安全性が低下するという問題があるが、この点についての考慮はされていない。

【0005】一般に、格納されるデータの内、相対的に重要度の高いデータをどの冗長群に格納するかを予め決

めて運用することが、ユーザ又は保守員等にとって可能な時においては、相対的に重要なデータの格納される冗長群に対して優先的に予備ディスクドライブをバックアップ用に使用した方が、一定数の予備ディスクドライブ数に対して効率良くデータの保全性をより高くすることが出来る。しかし、予め各冗長群に対して全て固定的に予備ディスクドライブを割当ててしまうと、予備ディスクドライブの効率的な使用が出来なくなる。

【0006】よって、バックアップ処理を自動的に開始する装置において、ユーザが相対的に重要度が高いと判断するデータの格納された冗長群に対し予備ディスクドライブを優先的に割当てるということと、与えられた予備ディスクドライブを効率良く使用することにより必要な信頼性を確保するために要するコストを低く抑えることの両立を実現することが、本発明の課題となる。

【0007】

【課題を解決するための手段】上記課題を解決するため、本発明は、複数の独立なディスクドライブを制御し、中央処理装置と、前記複数のディスクドライブとの間におけるデータの授受を制御し、前記複数のディスクドライブの内少なくとも1つに対して、前記データを配置し記憶させ、前記中央処理装置からの前記複数のディスクドライブに対するアクセス要求を前記データ配置に合わせて処理する制御論理を含み、前記複数のディスクドライブのいずれかに発生した障害を認識し、それがどのディスクドライブに発生した障害なのかを認識し、前記データの授受を中断すること無しに当該ディスクドライブ上に記憶されているデータを前記複数のディスクドライブの内の予備ディスクドライブへバックアップを行う記憶制御装置であり、前記複数のディスクドライブの内少なくとも1つを、予備ディスクドライブとして保持し、バックアップ処理の際にどのディスクドライブに対しても使用する共用の予備ディスクドライブとするか、あるいはバックアップ処理の際に特定の1つ以上のディスクドライブに対してのみ使用する専用の予備ディスクドライブとするか、という属性を予め設定する手段と、バックアップ処理を開始する際に、個々の前記予備ディスクドライブの属性の設定内容を参照して前記予備ディスクドライブの内の1つをバックアップ処理用に選択する選択手段を備えるようにしている。

【0008】また、前記選択手段は、障害の発生したディスクドライブに対して前記予備ディスクドライブの内の1つを選択するにあたって、前記共用の予備ディスクドライブの内未使用のものが1つ以上あり、かつ前記障害の発生したディスクドライブのための専用の予備ディスクドライブの内未使用のものが1つ以上あるとき、前記共用の予備ディスクドライブの内未使用のものの中の1つを選択するようにしている。

【0009】また、前記データを複数のセグメントに分け、各セグメントを前記複数のディスクドライブの内、

異なるディスクドライブに格納してデータを配置し、前記中央処理装置からの前記記憶装置に対するアクセス要求を前記データ配置に合わせて処理する手段を更に備えるようにしている。

【0010】また、前記データを複数のセグメントに分け、更にそれらを元に冗長データを作成し、前記セグメントと冗長データを前記複数のディスクドライブの内、異なるディスクドライブに格納してデータを配置し、前記中央処理装置からの前記記憶装置に対するアクセス要求を前記データ配置に合わせて処理する手段を更に備えるようにしている。

【0011】さらに、前記各セグメントと冗長データを格納した複数のディスクドライブの内の1台のディスクドライブに障害が発生し、該障害が発生したディスクドライブに対するバックアップ処理をするとき、障害の発生したディスクドライブ以外のディスクドライブに格納されたデータを用いて、障害の発生したディスクドライブ上に格納されていたデータの再構成をすることにより、バックアップ処理を行う手段を更に備えるようにしている。

【0012】さらに、前記バックアップ処理において、障害が発生したディスクドライブの障害の程度により、前記のデータの再構成に依るバックアップ処理を行うか、単純に前記障害の発生したディスクドライブ上に格納されているデータを複写することによるバックアップ処理を行うかを切替える手段を更に備えることを特徴とする記憶制御装置。

【0013】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。図1は、本発明の一実施の形態であるディスクサブシステムを含む情報処理システムの構成の概略構成図である。中央処理装置10と、記憶制御部30及び記憶装置部40を含む記憶制御装置20とで構成されている。中央処理装置10と記憶制御部30とは、情報伝送線60を介して結合され、記憶制御部30と記憶装置部40とは情報伝送線70-1及び情報伝送線70-2で結合されている。記憶装置部40では、ディスクドライブ41-1～41-8が情報伝送線70-1で記憶制御部30内の制御部31に接続され、ディスクドライブ41-9～41-16が情報伝送線70-2で記憶制御部30内の制御部31に接続されている。記憶制御部30は、中央処理装置10と記憶装置部40との間におけるデータの授受を制御する制御部31と、サービスプロセッサ36と、共有メモリ32で構成される。

【0014】共有メモリ32は、制御部31及びサービスプロセッサ36によって参照され、予備ディスクドライブ管理情報テーブル34と、記憶装置部40内のディスクドライブ41-1～41-16へのデータの配置を一意に特定するためのデータ配置情報テーブル33と、

障害統計情報テーブル35とを含む。障害統計情報テーブル35には、個々のディスクドライブ41-1～41-16の障害に関する統計情報を格納する。制御部31は、予備ディスクドライブ選択手段102を含む。

【0015】サービスプロセッサ36は予備ディスクドライブ属性設定手段101を含む。サービスプロセッサ36は、保守端末50を介して外部の図示しない操作者すなわちユーザ又は保守員によって制御され、データ配置情報テーブル33及び予備ディスクドライブ管理情報テーブル34に必要な情報の設定を行う。すなわち、データ配置情報テーブル33に対しては、個々の冗長群及び非冗長群の構成の設定、すなわち、ディスクドライブ41-1～41-16の内どれをどの冗長群及び非冗長群の何番目の構成ディスクドライブに指定するかの設定と、配置法、例えば、データ用ドライブ6台とパリティ用ディスクドライブ1台から成るRAID5とするか、データ用ドライブ1台とパリティ用ディスクドライブ1台のRAID1とするか、あるいは1台で独立したディスクドライブ群となるRAID0とするか、等の配置法の種別についての設定と、各冗長群及び非冗長群に対し、それぞれ幾つの論理デバイスをどのような順番で割り付けるかの設定と、を行う。予備ディスクドライブ管理情報テーブル34に対しては、同じく操作者により保守端末を介して予備ディスクドライブ属性設定手段101を用いて、ディスクドライブ41-1～41-16のどれを予備ディスクドライブに設定するか、また、個々の予備ディスクドライブを、どの冗長群または非冗長群に対しても使用する共用の予備ディスクドライブとするか、あるいは特定の冗長群または非冗長群に対してのみ使用する専用の予備ディスクドライブとするか、それぞれの冗長群または非冗長群に対して専用に設定するか、の設定を行う。

【0016】中央処理装置10から論理デバイスへの読み出し又は書き込み要求があるとき、そのデバイスアクセスアドレスと、記憶装置部40内の個々のディスクドライブ41-1～41-16及びそのディスクドライブ内での記憶領域との対応を、制御部31がデータ配置情報テーブル33と、必要ならば予備ディスクドライブ管理情報テーブル34と、を用いて算出し、制御部31が各ディスクドライブを制御する。予備ディスクドライブ管理情報テーブル34を参照するのは、バックアップ処理中及び処理後は本来のディスクドライブではなく、バックアップ先ディスクドライブをアクセスする必要が生じ得るからである。

【0017】操作者は、前述の論理デバイスの割当てと、予備ディスクドライブ割当てとを組み合わせ、特定の少なくとも1つ以上の論理デバイスを含むディスクドライブ群に予備ディスクドライブを専用に割当て、更に中央処理装置から当該の論理デバイスに対し、相対的に重要なデータを配置することにより、当該データに対

し、予備ディスクドライブを優先的に割当てることが可能となる。

【0018】各ディスクドライブ41-1～41-16への制御時におけるエラーの発生及び障害の報告は、制御部31によって選別・整理されて、共有メモリ32内の障害統計情報テーブル35に格納・更新される。更新時、統計情報から制御部31がバックアップ処理が必要であると判断した場合に、予備ディスクドライブ選択手段102を用いて自動的に予備ディスクドライブの内1つを選択し、それをバックアップ先ディスクドライブにして、障害の発生したディスクドライブ上のデータのバックアップ処理を自動的に開始する。

【0019】バックアップの方法としては、単純な複写による方法か、あるいは冗長性を用いて障害ディスクドライブ上のデータを再構成してバックアップ先ディスクドライブに書き込む方法か、のいずれかである。このバックアップ処理の開始の判断において、特定種類の障害の累計が閾値を越えたと判断された場合には、データの喪失に至る前に、予防的に冗長性を用いない方法により、すなわち単純な複写による方法により、予備ディスクドライブへのバックアップを行ってもよい。この方法は、その時点で当該のディスクドライブ群に冗長性がある場合及び冗長性が無い場合のいずれにおいても、データの保全性を維持するのに有効な方法である。また、バックアップ処理の開始の判断において、特定種類の障害の累計が別の閾値を越えたと判断された場合には、その時点で当該のディスクドライブ群に冗長性がある場合において、冗長性を用いて当該障害発生ディスクドライブ上のデータを再構成することにより、予備ディスクドライブへのバックアップを行ってもよい。

【0020】予備ディスクドライブの選択において、障害の発生したディスクドライブ群にとって、専用の予備ディスクドライブと、共用の予備ディスクドライブの両方に未使用の予備ディスクドライブがある場合は、共用または専用の予備ディスクドライブの内からいずれか1つをバックアップ先ディスクドライブとして選択してもよく、また、優先して共用の予備ディスクドライブの内から1つをバックアップ先ディスクドライブに選択し、共用の予備ディスクドライブで未使用のものが1つも無くなった状態のときにのみ、専用の予備ディスクドライブのうちの1つをバックアップ先ディスクに選択するようにしてもよい。

【0021】図2は、本発明の一実施の形態であるディスクサブシステムを含む情報処理システムのハード構成図である。図1の原理説明図における制御部は、図2において、ホストアダプタ231-1～231-2、ディスクアダプタ233-1～233-4、キャッシュメモリ232-1～232-2、データ転送バス237-1～237-2、記憶制御部内通信バス236、とを含む部位に該当する。記憶装置部内のディスクドライブ24

2-1～242-64はSCSIディスクドライブである。各ディスクアダプタ233-1～233-4はディスクドライブボックス241-1～241-2の内、特定のディスクドライブボックスしか制御しないが、個々のディスクドライブと制御部内のディスクアダプタ233-1～233-4とを結ぶ情報伝送線270-1～270-16は必ず2系統有り、耐障害性を高めているとともに、記憶制御部230に対し、応答性を高めている。中央処理装置210からのデータの書き込み要求に対し、2つのキャッシュメモリ232-1～232-2に2重化してバッファリングし、中央処理装置210からのデータ形式である可変長形式から、SCSIディスクドライブ用の固定長形式に変換して、ディスクドライブに書き込む。よって、データは必ずキャッシュを介してアクセスされるため、データキャッシュメモリを不揮発化（96時間）し、また前述のようにデータを2重化することにより、書き込み要求データの喪失を防止する。中央処理装置210との間のデータの転送は、複数の情報伝送線260-1～260-8が用意されていて、耐障害性を高めるとともに、応答性を高めている。共有メモリ234-1～234-2は2重化されており、データ配置情報メモリなどが失われることを防止する。データ転送バス237-1～237-2は2重化されており、耐障害性と応答性を高めている。ホストアダプタ231-1～231-2、ディスクアダプタ233-1～233-4、はそれぞれが並列に動作可能であり、応答性を高めている。このように、中央処理装置210とディスクドライブ242-1～242-64との間でのデータの制御において、情報伝送線260-1～260-8、ホストアダプタ231-1～231-2、ディスクアダプタ233-1～233-4、データ転送バス237-1～237-2、キャッシュメモリ232-1～232-2、情報伝送線270-1～270-16、が全て冗長化されている。

【0022】図4は、記憶制御装置220の概観図を示す。幅約1.3m、奥行き約0.8m、高さ約1.8mの大きさである。装置には2系統の電源入力部があり、装置外での電源障害に対する耐性を高めている。記憶装置部230側のディスクドライブ電源供給単位部420-1～420-8の内、前面の一番下にあるディスクドライブ電源供給単位部420-1について、更に詳しく図5に示してある。これは最大8つのディスクドライブ242-1～242-8を含み、また、入力電源を直流に変換して、各部に電源を供給する2つの直流電源供給部421-1～421-2はそれぞれ冗長構成であって、どちらか一方が動作すれば、ディスクドライブ242-1～242-8は稼働可能である。ディスクドライブ電源供給単位部420-1～420-4で図2のディスクドライブボックス241-1を構成し、ディスクドライブ電源供給単位部420-5～420-8でディス

クドライブボックス241-2を構成する。

【0023】図6及び図7には、記憶制御部220の前面から見た様子と後面から見た様子を示している。図7において、論理部ボックスにはホストアダプタ231-1～231-2、ディスクアダプタ233-1～233-4、キャッシュメモリ232-1～232-2、共有メモリ234-1～234-2を実装する。直流電源集約ボックス322及び複数の直流電源供給部323-1～323-6はやはり2系統の冗長構成となっており、それぞれに対し、ブレーカボックス321-1～321-2を備えている。以上の2重化、及び多重化されている部位のうち、共有メモリ234-1～234-2、キャッシュメモリ232-1～232-2、ホストアダプタ231-1～231-2、ディスクアダプタ233-1～233-4、ディスクドライブ242-1～242-64、直流電源供給部323-1～323-6及び421-1～421-16、及び直流電源集約ボックス322内の個々の直流電源供給部は、活線挿抜にて増減設及び交換が可能となっており、中央処理装置との間のデータの授受を続けながら、障害発生時の保守に加えて、定期保守、ハード構成の増減設が可能である。

【0024】次に、幾つかの例を示して、図1の原理説明図において、中央処理装置10からのアクセス要求の処理及びバックアップ処理時の予備ディスクドライブの選択を説明する。図8は、記憶装置部40内の個々のディスクドライブの論理的な設定の一例を示す。ディスクドライブ41-1～41-16が、操作者による初期設定によって、予備ディスクドライブ4個（ディスクドライブ41-13～41-16）が設定され、残りのディスクドライブ41-1～41-12は、それぞれディスクドライブ41-1～41-3、41-4～41-6、41-7～41-9、41-10～41-12から成るディスクドライブ群90-1～90-4に論理的に分けられている。この例では、実装可能なディスクドライブ全て実装し、それら全てを予備ディスクドライブ又はデータ及び冗長データを格納するディスクドライブとして設定している。それぞれのディスクドライブ群ではデータと冗長データが容量においてそれぞれディスクドライブ2台分と1台分を占める、ディスクドライブ3台から構成されるRAID5の冗長ディスクドライブ群に設定されている。予備ディスクドライブの内、ディスクドライブ41-13はディスクドライブ群90-1専用、ディスクドライブ41-14はディスクドライブ群90-2専用、ディスクドライブ41-15とディスクドライブ41-16は共用の予備ディスクドライブとして設定されている。

【0025】図8のような構成においては、ディスクドライブ群90-1及び90-2は、予備ディスクドライブの割当ての観点において、少なくとも1回、最大で3回までのバックアップ処理を行うことが可能である。す

なわち、1台の予備ディスクドライブは1度使用されればそれ以上の使用はできない（但し、障害の発生したディスクドライブの障害回復が行われ、予備ディスクドライブに格納されているデータが障害回復されたディスクドライブに逆複写された場合には、その予備ディスクドライブを再び使用することはできる）。そのため、共用の予備ディスクドライブを最大2回使用でき、専用の予備ディスクドライブを1回使用できるため、都合、最大で3回までのバックアップ処理を行うことが可能であり、2台の共用の予備ディスクドライブが他のディスクドライブ群で使用されてる場合には、少なくとも1回のバックアップ処理を行うことが可能である。一方、ディスクドライブ群90-3及び90-4は、専用の予備ディスクドライブが設定されていないため、最大で2回までの共用の予備ディスクドライブによるバックアップ処理が可能だが、別のバックアップ処理が事前に始まっている時には、一回も行うことが出来ないことがあり得る。

【0026】図8のような設定における、ある時点での論理的な状態を図9に示す。図12の表1は、この時の予備ディスクドライブ管理情報テーブル34上の内容の一部を示す。表1においての各項目は、以下のような意味である。予備ディスクドライブ位置番号は、予備ディスクドライブの記憶装置部40内における各ディスクドライブの配置位置を示す番号である。専用属性フラグは、それがONの時に、バックアップ処理の際に特定ディスクドライブ群に対して専用使用する属性に設定していることを表している。使用状態は、単純バックアップ中か、バックアップ済か、冗長性を用いたバックアップ中か、逆複写中か、未使用か、を表わす。対応ディスクドライブ群通番は、専用属性フラグがONの時にはそのディスクドライブ群通番の指し示すディスクドライブ群に対して専用使用するよう設定されていることとし、専用属性フラグがOFFかつ使用状態が未使用状態以外の時には、それがどのディスクドライブ群通番の指し示すディスクドライブ群に対して使用しているかを表す。01はディスクドライブ群90-1に、02はディスクドライブ群90-2に、03はディスクドライブ群90-3に対応する。処理対象ディスクドライブの相対位置は、バックアップ処理対象ディスクドライブの、そのディスクドライブ群内における相対位置を表す。例えば、表1の最上欄の2はディスクドライブ群90-1のディスクドライブ41-2に対応する。

【0027】図3は、予備ディスクドライブ管理情報テーブル34を参照して、予備ディスクドライブ選択手段102が予備ディスクドライブを選択するときのフローチャートを示している。図9において、×印の付いているディスクドライブにはバックアップ処理が必要な障害が発生していることを示し、矢印はそのバックアップ先ディスクドライブを指し示している。×印の付いている

ディスクドライブの内、ディスクドライブ41-12が最も最近においてバックアップ処理必要と判断された、障害のあるディスクドライブであるとする。このとき、共用の予備ディスクドライブに未使用のものが1つも無いために、ディスクドライブ41-12のバックアップ処理は行わない。すなわち、図3のフローチャートにおいて、ステップ1001「バックアップ処理要求発生」、ステップ1002「専用でない予備ディスクドライブの内未使用のものがあある」に対しNO、ステップ1004「当該ディスクドライブ群に専用の予備ディスクドライブが設定されている」に対しNO、ステップ1007「バックアップ処理不可時処理へ」の流れとなる。よって、以後ディスクドライブ41-12へのアクセスは、ディスクドライブ群の冗長性を用いて行われる。また、以後ディスクドライブのいずれかにおいて、バックアップ処理が必要な障害が発生した際には、最後の未使用予備ディスクであるディスクドライブ41-14が専用予備ディスクドライブとして設定されている、ディスクドライブ群90-2を構成するディスクドライブ41-4~41-6のいずれかについてでなければ、バックアップ処理は行われない。このバックアップ処理が動作するときは、図3のフローチャートにおいて、ステップ1001「バックアップ処理要求発生」、ステップ1002「専用でない予備ディスクドライブの内未使用のものがあある」に対しNO、ステップ1004「当該ディスクドライブ群に専用の予備ディスクドライブが設定されている」に対してYES、ステップ1005「専用の予備ディスクドライブの内未使用のものがあある」に対しYES、ステップ1006「専用の予備ディスクドライブで、かつ未使用のものから1つをバックアップ処理用に指定」、ステップ1008「バックアップ処理開始へ」の流れとなる。ただし、保守員により予備ディスクドライブの増設が行われた場合や、バックアップ処理を終えた障害ディスクの交換又は修理が行われてバックアップ先ディスクドライブの内容を逆複写する処理が行われた場合は、いずれのディスクドライブ群も再びバックアップ処理が可能となり得る。

【0028】図10は、記憶装置部40内の個々のディスクドライブの論理的な設定の別の一例を示す。ディスクドライブ41-1~41-16が、操作者による初期設定によって、予備ディスクドライブ4個（ディスクドライブ41-13~41-16）が設定され、残りのディスクドライブ41-1~41-12の内、ディスクドライブ41-1~41-10は、それぞれディスクドライブ41-1~41-3、41-4~41-6、41-7~41-8、41-9、41-10から成るディスクドライブ群90-5~90-9に論理的に分けられている。この例では、実装可能なディスクドライブ数の上限より2台少く実装され、実装分については全てを予備ディスクドライブ又はデータ及び冗長データを格納するデ

ィスクドライブとして設定している。ディスクドライブ群90-5、90-6ではデータと冗長データが容量においてそれぞれディスクドライブ2台分と1台分を占める、ディスクドライブ3台から構成されるRAID5の冗長ディスクドライブ群に設定されている。ディスクドライブ群90-7ではデータと冗長データが容量においてそれぞれディスクドライブ1台分ずつを占める、ディスクドライブ2台から構成されるRAID1の冗長ディスクドライブ群に設定されている。ディスクドライブ群90-8、90-9は、それぞれ1台のディスクドライブのみ独立したディスクドライブ群を構成する、RAID0の非冗長ディスクドライブ群に設定されている。予備ディスクドライブの内、ディスクドライブ41-13はディスクドライブ群90-5専用、ディスクドライブ41-14はディスクドライブ群90-6専用設定されており、ディスクドライブ41-15、ディスクドライブ41-16は共用の予備ディスクドライブとして設定されている。

【0029】図10のような構成においては、ディスクドライブ群90-5及び90-6は、予備ディスクドライブの割当ての観点において、少なくとも1回、最大で3回までのバックアップ処理を行うことが可能であり、一方、ディスクドライブ群90-7及び90-8、90-9は、最大で2回までのバックアップ処理が可能だが、別のバックアップ処理が事前に始まっている時には、一回も行うことが出来ないことがあり得る。

【0030】図10のような設定における、ある時点での論理的な状態を図11に示す。図12の表2は、この時の予備ディスクドライブ管理情報テーブル34上の内容の一部を示す。表2において、対応ディスクドライブ群通番の01はディスクドライブ群90-5に、02は90-6に、04は90-8に対応する。

【0031】図11において、×印の付いているディスクドライブにはバックアップ処理が必要な障害が発生していることを示し、矢印はそのバックアップ先ディスクドライブを指し示している。×印の付いているディスクドライブの内、ディスクドライブ41-10が最も最近においてバックアップ処理必要と判断された、障害のあるディスクドライブであるとする。このとき、共用の予備ディスクドライブに未使用のものが1つも無いために、ディスクドライブ41-10のバックアップ処理は行わない。すなわち、図3のフローチャートにおいて、ステップ1001「バックアップ処理要求発生」、ステップ1002「専用でない予備ディスクドライブの内未使用のものがあある」に対しNO、ステップ1004「当該ディスクドライブ群に専用の予備ディスクドライブが設定されている」に対してNO、ステップ1007「バックアップ処理不可時処理へ」の流れとなる。

【0032】ディスクドライブ41-10は非冗長なディスクドライブ群90-9に属しているので、バックア

ップ処理を必要とする判断が、予防的なバックアップを目的としているなら、以後ディスクドライブ41-10へのアクセスは、一部のデータが喪失する可能性が高い状態のままで行われる。また、以後ディスクドライブのいずれかにおいて、バックアップ処理が必要な障害が発生した際には、最後の未使用予備ディスクであるディスクドライブ41-13が専用予備ディスクドライブとして設定されている、ディスクドライブ群90-5を構成するディスクドライブ41-1~41-3、41-15のいずれかについてでなければ、バックアップ処理は行われない。すなわち、41-3に対するバックアップ処理が正常に終了しているならば、41-1、41-2、41-15がディスクドライブ群90-5を構成しており、一方、正常に終了する前ならば、41-1、41-2、41-3がディスクドライブ群90-5を構成しているので、41-13を用いたバックアップ処理は、前者において、41-1、41-2、41-15について可能であり、後者において、41-1、41-2、41-3について可能である。このように、ディスクドライブ41-3についてディスクドライブ41-15へのバックアップ処理が終了しているならば、ディスクドライブ41-15において障害が発生したときにも、ディスクドライブ41-15についてのバックアップ処理が可能である。

【0033】ディスクドライブ群90-5についてのバックアップ処理が動作するときは、図3のフローチャートにおいて、ステップ1001「バックアップ処理要求発生」、ステップ1002「専用でない予備ディスクドライブの内未使用のものがある」に対しNO、ステップ1004「当該ディスクドライブ群に専用の予備ディスクドライブが設定されている」に対してYES、ステップ1005「専用の予備ディスクドライブの内未使用のものがある」に対しYES、ステップ1006「専用の予備ディスクドライブで、かつ未使用のものから1つをバックアップ処理用に指定」、ステップ1008「バックアップ処理開始へ」の流れとなる。ただし、保守員により予備ディスクドライブの増設が行われた場合や、バックアップ処理を終えた障害ディスクの交換又は修理が行われてバックアップ先ディスクドライブの内容を逆複製する処理が行われた場合は、いずれのディスクドライブ群も再びバックアップ処理が可能となり得る。

【0034】以上の2つの構成例の後者において、非冗長なディスクドライブ群の例として単一のディスクドライブのみで構成されているものについて説明したが、これは複数のディスクドライブから成る非冗長群、例えばパリティデータを持たないRAID5などでもよい。

【0035】このように、本実施の形態によれば、未使用予備ディスクドライブが全体として数に余裕があるうちは全てのディスクドライブ群に対し等しくバックアップ処理を行い、一方、全体として数に余裕が無くなって

きた時は、予め専用の予備ディスクドライブが設定されていたディスクドライブ群に対し優先的にバックアップ処理を行う。すなわち、未使用予備ディスクドライブが全体として数に余裕があるうちは与えられた複数の予備ディスクドライブを効率良く使用することが出来、一方、全体として数に余裕が無くなってきた時は、相対的に重要度の高いデータが格納されていると予め操作者が判断して予備ディスクドライブを専用を設定しておいたディスクドライブ群に対して、優先的に予備ディスクドライブをバックアップディスク用に使用しそのディスクドライブ群上のデータ保全性を高める。これらにより、要求されている信頼性を確保するために要するコストを最小化している。

【0036】以上、本発明を実施の形態に基づき具体的に説明したが、本発明は前記実施例に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能である。

【0037】

【発明の効果】本発明の記憶制御装置によれば、ディスクドライブの障害発生時に、中央処理装置とのデータの授受を続行しながら自動的にバックアップ処理を開始する装置において、相対的に重要度の高いデータの格納されたディスクドライブに対し、ユーザ及び保守員等の設定操作により予備ディスクドライブを優先的に使用することを可能としながら、与えられた予備ディスクドライブを効率良く使用することにより必要な信頼性を確保するために要するコストを最小化するという効果が得られる。

【図面の簡単な説明】

【図1】本発明の一実施の形態である記憶制御装置を含む情報処理システムの概略構成を示す図である。

【図2】本発明の一実施の形態である記憶制御装置を含む情報処理システムのハード構成を示す図である。

【図3】本発明の一実施の形態である記憶制御装置において、バックアップ処理が必要になった際の処理を示すフローチャートである。

【図4】本発明の一実施の形態である記憶制御装置の概観図である。

【図5】本発明の一実施の形態である記憶制御装置のディスクドライブ電源供給単位部の概観図である。

【図6】本発明の一実施の形態である記憶制御装置を構成する記憶制御部の前面よりの概観図である。

【図7】本発明の一実施の形態である記憶制御装置を構成する記憶制御部の後面よりの概観図である。

【図8】本発明の一実施の形態である記憶制御装置が制御する記憶装置内の個々のディスクドライブに関する論理的な設定の一例を示す図である。

【図9】図8の論理的な設定において、バックアップ処理が行われている、ある特定時点での論理的な状態を示す図である。

【図10】本発明の一実施の形態である記憶制御装置が制御する記憶装置内の個々のディスクドライブに関する論理的な設定の別の一例を示す図である。

【図11】図10の論理的な設定において、バックアップ処理が行われている、ある特定時点での論理的な状態を示す図である。

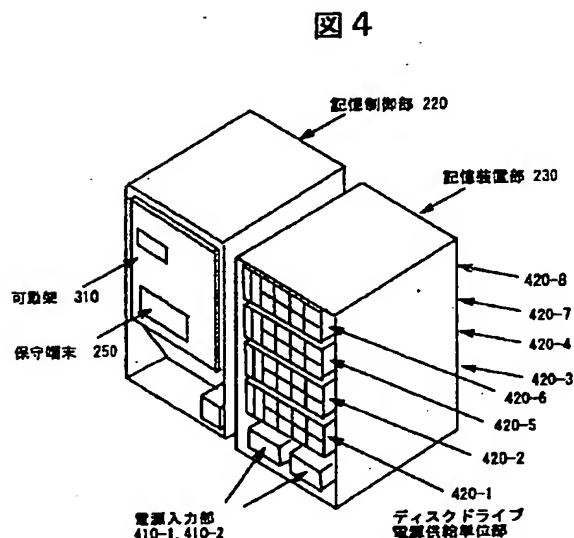
【図12】図9と図11の状態における、予備ディスクドライブ管理情報の一例を夫々表として示す図である。

【符号の説明】

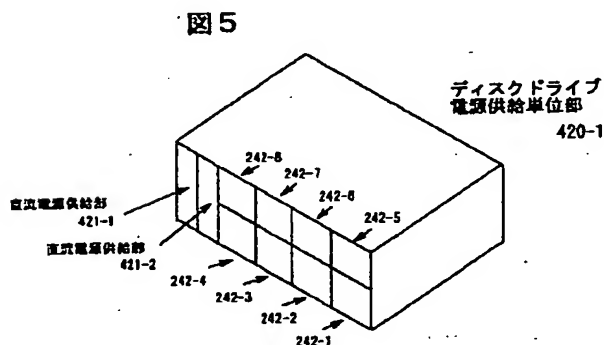
10 中央処理装置
20 記憶制御装置
30 記憶制御部
31 制御部
32 共有メモリ
33 データ配置情報テーブル
34 予備ディスクドライブ管理情報テーブル
35 障害統計情報テーブル
36 サービスプロセッサ
40 記憶装置部
41-1～41-16 ディスクドライブ
50 保守端末
60 情報伝送線
70-1～70-2 情報伝送線
90-1～90-9 ディスクドライブ群
101 予備ディスクドライブ属性設定手段
102 予備ディスクドライブ選択手段

210 中央処理装置
220 記憶制御装置
230 記憶制御部
231-1～231-2 ホストアダプタ
232-1～232-2 キャッシュメモリ
233-1～233-4 ディスクアダプタ
234-1～234-2 共有メモリ
235 サービスプロセッサ
236 制御部内通信バス
237-1～237-2 データ転送バス
240 記憶装置部
241-1～241-2 ディスクドライブボックス
242-1～242-64 ディスクドライブ
250 保守端末
260-1～260-8 情報伝送線
270-1～270-16 情報伝送線
310 可動架
321-1～321-2 ブレーカボックス
322 直流電源集約ボックス
323-1～323-6 直流電源供給部
324 増設用キャッシュボックス
325 論理部ボックス
410-1～410-2 電源入力部
420-1～420-8 ディスクドライブ電源供給単位部
421-1～421-16 直流電源供給部

【図4】

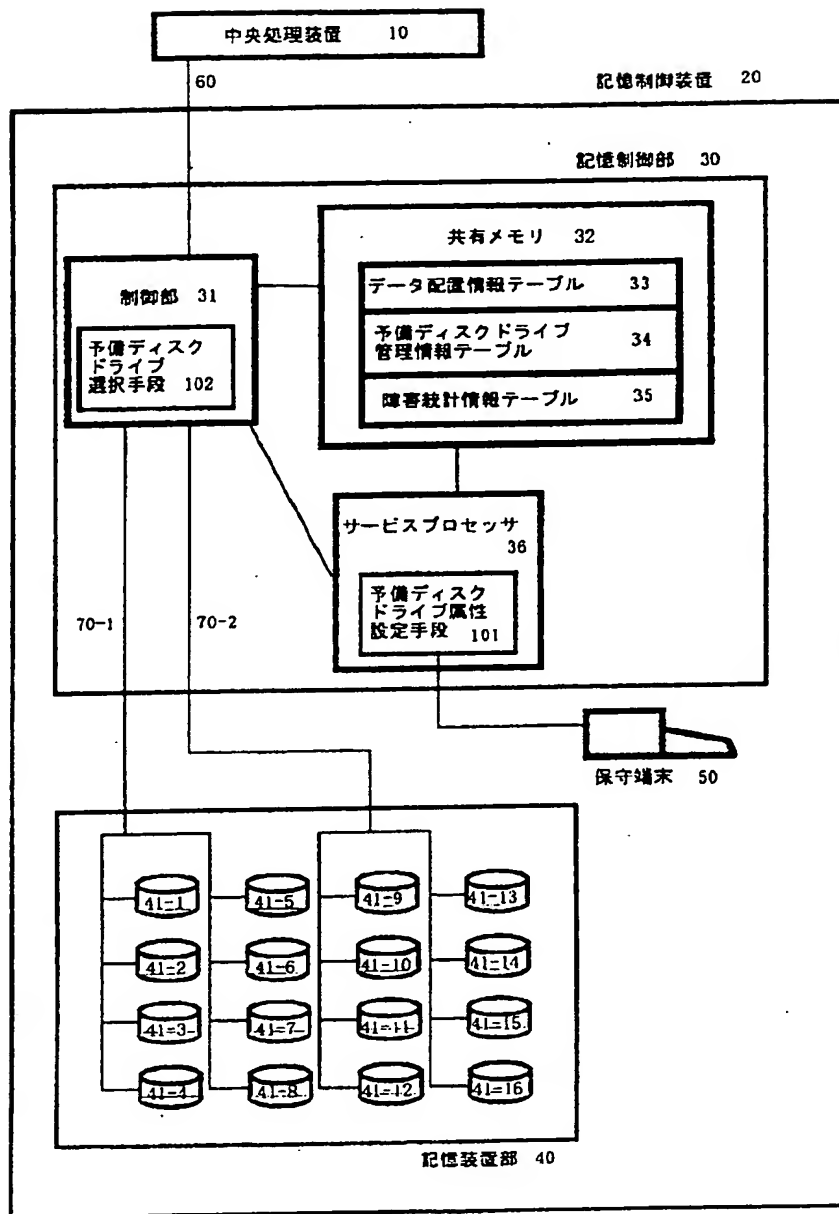


【図5】

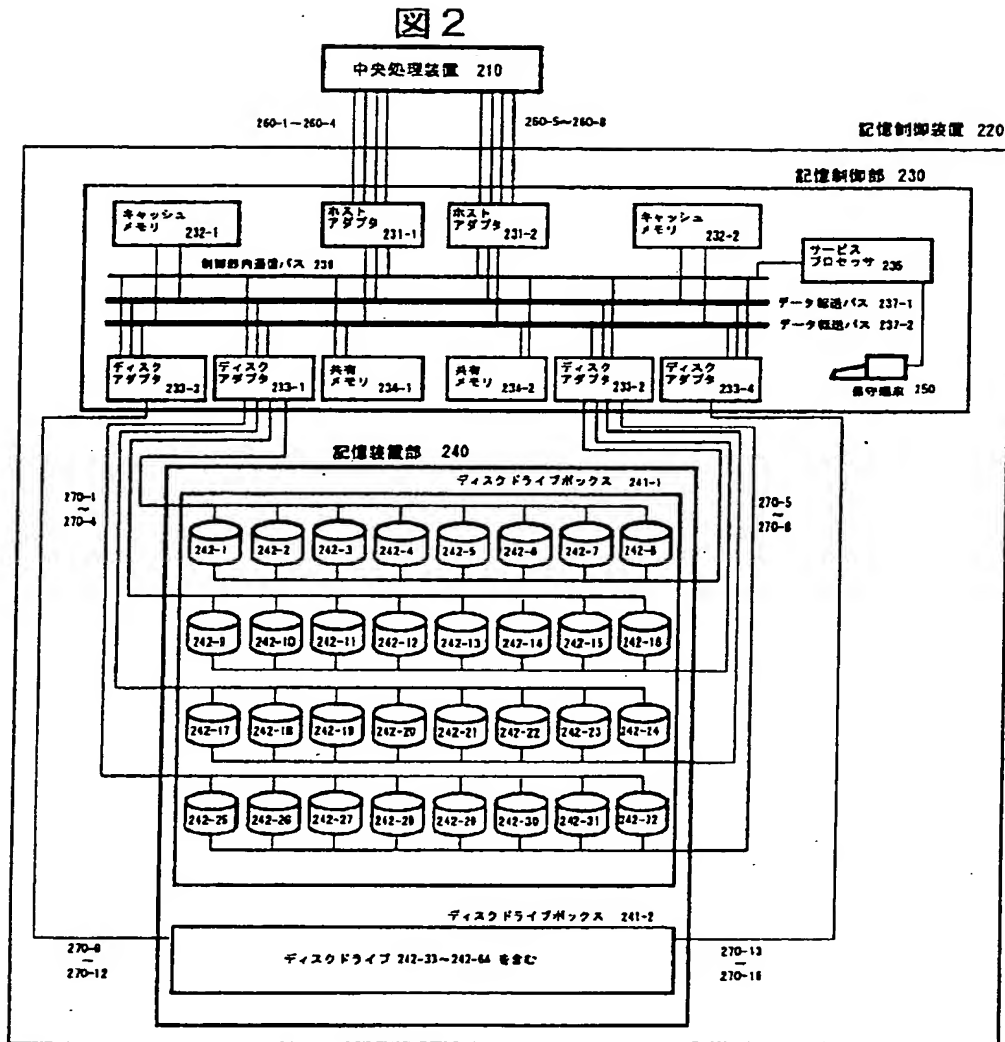


【図 1】

図 1

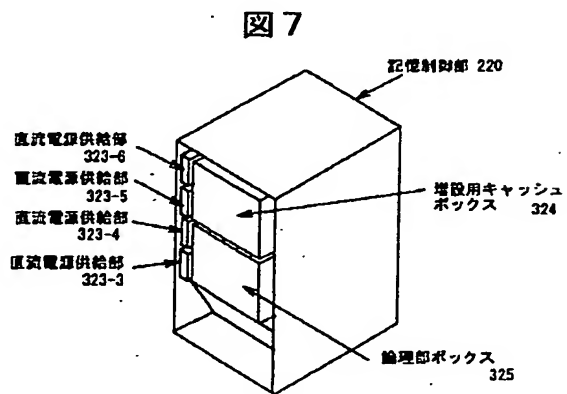
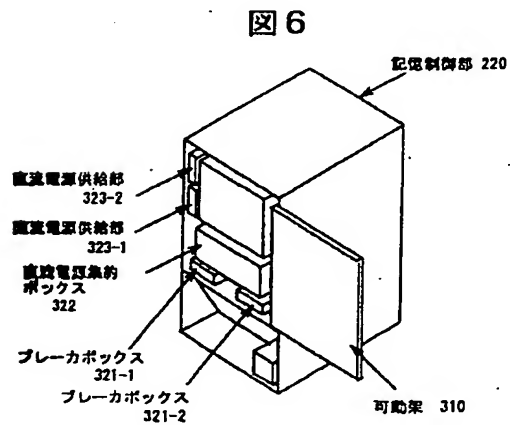


【図 2】



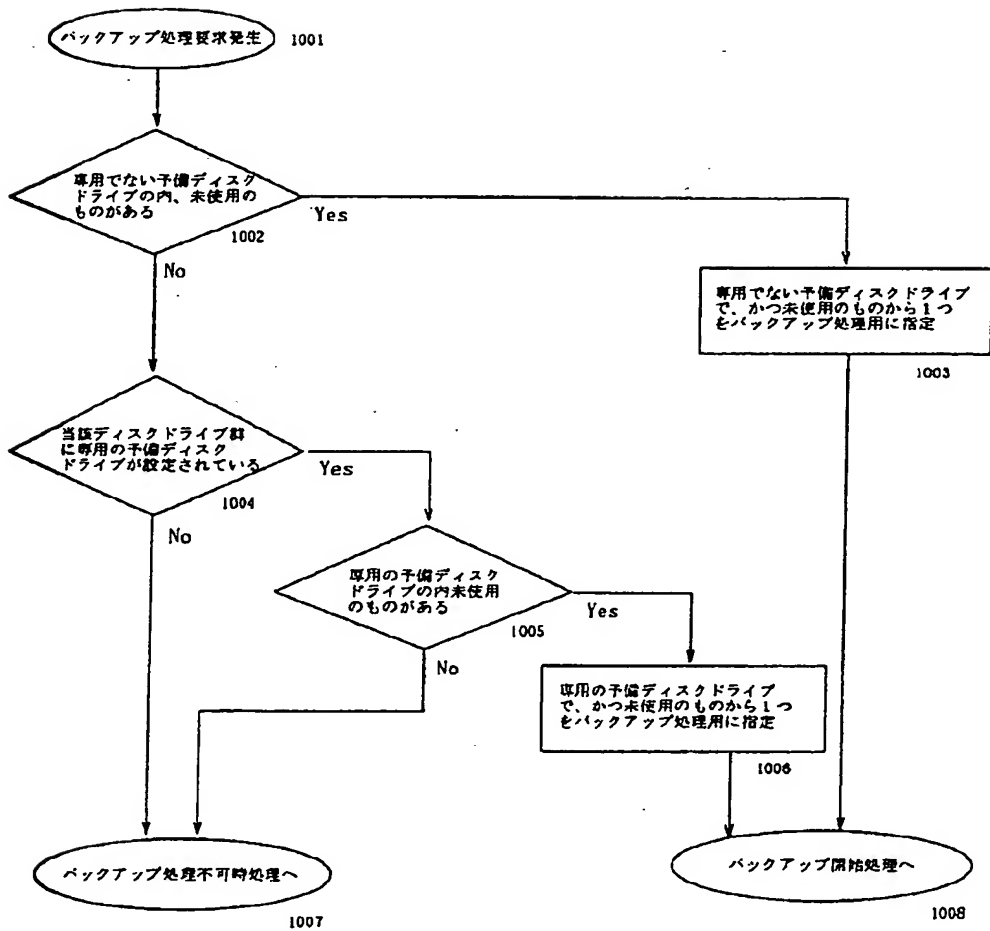
【図 6】

【図 7】



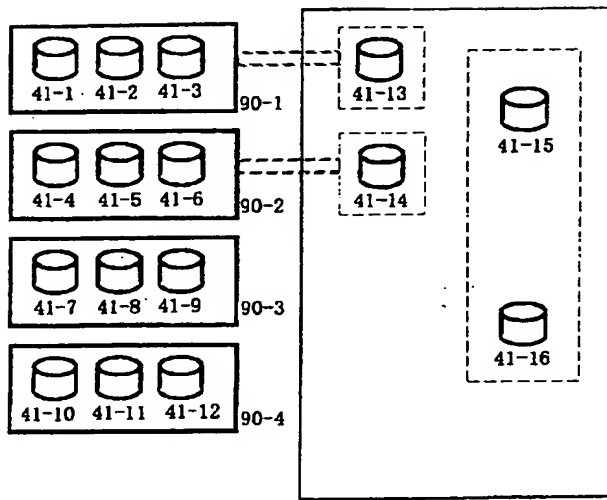
【図3】

図3



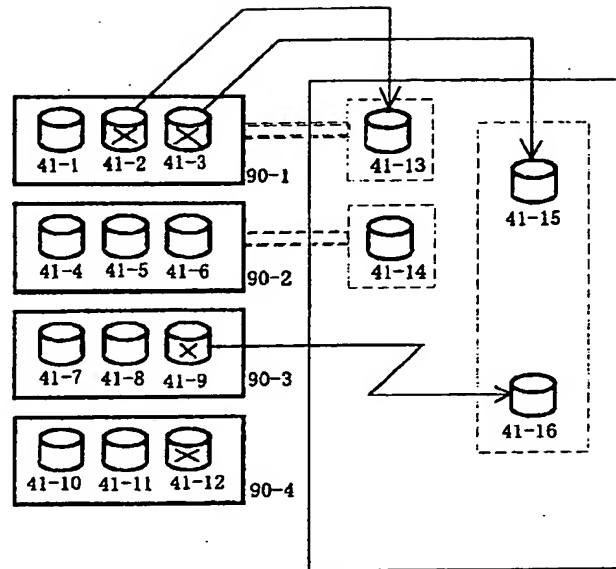
【図 8】

図 8



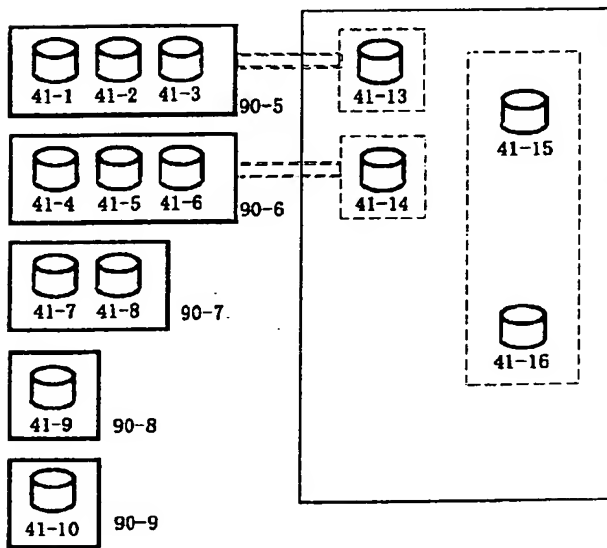
【図 9】

図 9



【図 10】

図 10



【図 11】

図 11

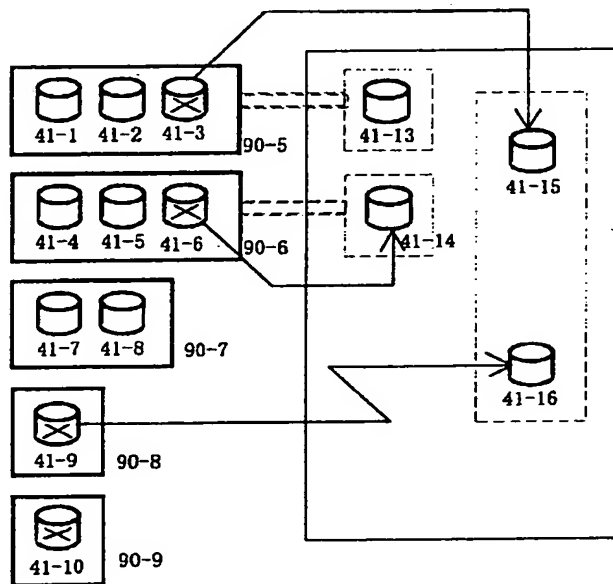


図12

表 1

予備ディスクドライブ位置番号	専用属性フラグ	使用状態	対応ディスクドライブ群通番	処理対象ディスクドライブ相対位置
13	ON	冗長性を用いてバックアップ中	01	2
14	ON	未使用	02	0
15	OFF	バックアップ済み	01	3
16	OFF	単純バックアップ中	03	3

表 2

予備ディスクドライブ位置番号	専用属性フラグ	使用状態	対応ディスクドライブ群通番	処理対象ディスクドライブ相対位置
13	ON	未使用	01	0
14	ON	冗長性を用いてバックアップ中	02	3
15	OFF	バックアップ済み	01	3
16	OFF	単純バックアップ中	04	1

フロントページの続き

(72)発明者 尾形 幹人

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内